

# Mercedes-Benz-Greener-Manufacturing

You are required to reduce the time that cars spend on the test bench. You will work with a dataset representing different permutations of features in a Mercedes-Benz car to predict the time it takes to pass testing. Optimal algorithms will contribute to faster testing, resulting in lower carbon dioxide emissions without reducing Daimler's standards.

## Step1: Import the required libraries

Step1.1: linear algebra

Step1.2: data processing

Step1.3: for dimensionality reduction

## Step2: Read the data from train.csv

Step2.1: let us understand the data

Step2.2: print few rows and see how the data looks like

## Step3: Collect the Y values into an array

Step3.1: separate the y from the data as we will use this to learn as the prediction output

## **Step4: Understand the data types we have**

Step4.1:iterate through all the columns which has X in the name of the column

## **Step5: Count the data in each of the columns**

## **Step6: Read the test.csv data**

Step6.1: remove columns ID and Y from the data as they are not used for learning

## **Step7: Check for null and unique values for test and train sets**

## **Step8: If for any column(s), the variance is equal to zero, then you need to remove those variable(s).**

Step8.1: Apply label encoder

## **Step9: Make sure the data is now changed into numericals**

## **Step10: Perform dimensionality reduction**

Step10.1: Linear dimensionality reduction using Singular Value Decomposition of the data to project it to a lower dimensional space.

## **Step11: Training using xgboost**

**Step12: Predict your test\_df values using xgboost**